DOCUMENT RESUME

ED 411 866 IR 056 693

AUTHOR Pollitt, A. Steven; Smith, Martin P.; Treglown, Mark; Braekevelt, Patrick

TITLE View-Based Searching Systems--Progress Towards Effective Disintermediation.

PUB DATE 1996-00-00

NOTE 15p.; In: Online Information 96. Proceedings of the International Online Information Meeting (20th, Olympia 2, London, England, United Kingdom, December 3-5, 1996); see IR 056 631.

PUB TYPE Reports - Evaluative (142) -- Speeches/Meeting Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.

DESCRIPTORS Access to Information; Biomedicine; Databases; Foreign Countries; Hypermedia; Information Dissemination; *Information Retrieval; *Information Seeking; Online Searching; *Online Systems; Reference Services; Relevance (Information Retrieval); Search Intermediaries; *Search Strategies; User Needs (Information); User Satisfaction (Information); *Users (Information)

IDENTIFIERS *Disintermediation; Europe; European Parliament; United Kingdom

ABSTRACT

This paper presents the background and then reports progress made in the development of two view-based searching systems--HIBROWSE for EMBASE, searching Europe's most important biomedical bibliographic database, and HIBROWSE for EPOQUE, improving access to the European Parliament's Online Query System. The HIBROWSE approach to searching promises to provide significantly more effective information retrieval for end-users than is possible through simple keyword, command line, forms-based or hypertext linking interaction. View-based searching makes extensive use of knowledge structures in the form of thesauri and classification schemes to provide linked browsable subject views onto databases. The result is a rich interface where queries can be satisfied by selective progressive refinement and expansion of mutually dependent views. The effect for the user is to significantly increase searching power without a commensurate increase in user effort, thereby reducing the reliance on intermediaries for sophisticated searching. (Contains 17 references.) (Author)

View-Based Searching Systems --- Progress Towards Effective Disintermediation

By:

A. Steven Pollitt

Martin P. Smith

Mark Treglown

Patrick Braekevelt

View-based searching systems — progress towards effective disintermediation

A. Steven Pollitt, Martin P. Smith, Mark Treglown and Patrick Braekevelt
The Centre for Database Access Research, University of Huddersfield, UK

Abstract: *This paper presents the background and then reports progress made in the development of two view-based searching systems — HIBROWSE for EMBASE, searching Europe's most important biomedical bibliographic database, and HIBROWSE for EPOQUE, improving access to the European Parliament's Online Query System. The HIBROWSE approach to searching promises to provide significantly more effective information retrieval for end-users than is possible through simple keyword, command line, forms-based or hypertext linking interaction. View-based searching makes extensive use of knowledge structures in the form of thesauri and classification schemes to provide linked browsable subject views onto databases. The result is a rich interface where queries can be satisfied by selective progressive refinement and expansion of mutually dependent views. The effect for the user is to significantly increase searching power without a commensurate increase in user effort, thereby reducing the reliance on intermediaries for sophisticated searching.*

Keywords: Disintermediation, view-based searching, end-users, user-interfaces, thesauri, classification, EMBASE, EMTREE, EPOQUE, EUROVOC, HIBROWSE

1. Introduction

'Disintermediation' first appears in the INSPEC database in the translated title of a conference paper presented in Milan in 1983 (Ref 6) — 'Microelectronics and teleinformatics as economic and social disintermediation tools. (Elimination of intermediaries between producers, distributors and consumers).' Subsequently it appears in only another eight records on the INSPEC database, yet the issue threatens to cause the most dramatic social changes that will affect the jobs people do since the industrial revolution. The word count for 'disintermediation' using the Alta Vista Web search engine was 665 on the 21 June 1996. The first ten sites mainly concern banking and electronic commerce, yet the third site in the list is 'Online Information 96: Call for Papers'.

Research and development at the University of Huddersfield has been directed towards enabling end-users of retrieval systems to make effective use of databases without the assistance of search intermediaries. All these efforts have made use of database specific thesauri where the user is able to apply the power of recognition and selection as the key characteristic of query specification. Expert system techniques (Ref 8) such as CANSEARCH (Refs 8,10) promised much in the early eighties but proved expensive to build and inferior in performance to the human intermediary. Automatic search statement generation from thesaurus-based menus (Refs 11,12) provided greater generality, more economic development and more powerful searching. This approach required the user to examine search statements describing sets of documents from combining search term selections and did not take full advantage of the user interface to assist the user in their search process. Amending the search statement was not straightforward and the general nature of the interaction had more affinity with batch rather than interactive processing.

This paper describes the most recent developments which evolved from the menu-based approaches in what has been termed 'view-based searching'. This approach was likened to the use of optical coincidence cards (Ref 13) because of the seemingly parallel nature of the processing, yet it offers considerably greater flexibility in the content and filtering options than could have been conceived of when these cards were employed over 30 years ago. The user is now provided with much more opportunity to examine the database and to apply powerful searching as we can simultaneously present several views and employ them to examine the contents of the database by refinement and expansion of different searching elements. Recently this work has been directed at the subjects of medicine and politics and law, through the development of HIBROWSE systems for EMBASE, the major European biomedical bibliographic database, and EPOQUE, the European Parliament's Online Query System.

It is not yet clear to what extent this approach liberates the user from knowledge of the workings of the retrieval system. The paradigm shift to searching through reciprocally refining views has moved the interaction between user and machine further towards the subject matter and away from the operation of the system. This causes the system designer difficulties in modelling tasks and human-computer dialogues (Ref 17).

2. The need for knowledge and skills in information searching

Education and training that enables an individual to play a specialist part in society has been a feature of our existence for thousands of years; no doubt as hunter gatherers we had those who were better at hunting than gathering and vice versa. The division of labour has sought to provide economic solutions to the production of goods and the provision of services. The case for intermediaries to undertake searches for information on behalf of researchers was made by Ranganathan in 1964 (Ref 15):

> 'The present rate of increase in the number of documents makes it wasteful to leave literature search in the hands of the research workers themselves. There is need to conserve the research potential by a division of labour. In this division of labour, literature search falls to the share of the library profession.'

This reflected the level of knowledge and skills required and the 'unproductive' research time that would be wasted in the days before online searching. Since the onset of computerised searching there have been efforts directed at making it possible for end-users to search for themselves, through approaches which reduce the levels of knowledge and skills required. These originally concentrated on the mundane but fundamental problems of connecting to a searching system but have subsequently moved onto the removal of the need to learn a command language, and now to tools which assist the user in expressing the concepts of queries.

A paper at the Online Information event in 1985 (Ref 9) advocated the pursuit of the 'What' rather than the 'How' in designing interfaces for the end-user in information retrieval, together with a separation of the logical from the physical, if we were to make the most of available technology. The ease of use was likened to travelling from A to B in a taxi, being concerned only with the destination rather than the route.

This question was revisited by David Nicholas in his paper at Online Information 95 (Ref 5). His study, comparing the searching of FT PROFILE by journalists and intermediaries at the Guardian newspaper, discovered that journalists use minimal searching features (only 52% of the commands available) yet they remained generally satisfied with their searching 'thanks largely to the fact they are better at sifting through data in pursuit of relevance.' The command-based interaction here can be said still to depend on the 'How' rather than the 'What' for effective interaction. There was a suggestion that there was also a division of searching where the more complex searching was delegated to intermediaries.

The success of the World Wide Web as a medium for end-user access is in some respects due to the minimal level of knowledge, skills and training required. The hypertext mode of interaction exploits our ability to recognise and all the user has to do is select. This interaction has completely done away with the 'How'. Unfortunately the following of hypertext links as currently provided is unlikely to bring effective information retrieval. More traditional keyword searching through the significant search engine developments, such as Alta Vista, has introduced different levels of complexity for searching that require additional skills and knowledge for effective use. Classified indexes are also being used to help the end-user. The introduction of organisation to any information collection, both gathering associated information together and facilitating browsing, has been an investment intended to save time and improve search performance on the part of the searcher. This is a natural response to the disorganised resource we would otherwise be faced with.

There are still those who would have us believe that natural language queries and ranked output offer the panacea for end-user searching. The size of result sets, and the potential for organising these according to different views, suggests a continuing role for classification as de Grolier stated some thirty years ago (Ref 3):

> 'We feared some years ago that classification was becoming useless, that the treatment of natural language texts by machine ... would replace classification. Classification and the classificationists would become something like the dinosaurs, killed by the progress of evolution. This has proved to be a complete fallacy. When you examine the new literature you find that more and more classification ... is considered as something quite essential in information retrieval ... It is quite evident that hierarchies, generally speaking, are something which cannot be avoided in an information retrieval system which is to be useful for the reader.'

The increased use of knowledge structures will be seen to become more important as they help in the building of user interfaces which minimise the skill but maximise the power available to the user to search databases.

3. A faceted organisation of knowledge

In the subject area of medicine the application of view-based searching relies on the professional knowledge of the end-user to progress a search. The use of facets in thesauri provides not only a level of organisation but the key to a form of processing which makes the task of information retrieval one of refinement or expansion of component facets, in the discovery of potentially relevant records.

The suggestion that a faceted classification scheme should be used as the basis of all methods of information retrieval was made by the Classification Research Group in 1955 (Ref 1). Forty years later the potential for faceted thesauri in the new century is described by Ron Davies (Ref 2) including the application for searching free text:

> 'It also suggests new ways to search full text databases, where only the words actually used in a document by its

author are searchable. If a thesaurus with a wide variety of entry terms is linked to a document database, it can provide a way of mapping the wide range of terms used in natural language to a single concept in the controlled vocabulary. A user entering a word in a search expression can have the search expanded to include all of the other possible words or expressions that are related in meaning automatically, thereby providing search results with greater recall.'

Assessing the usability of view-based searching system prototypes (Ref 16) forced a redesign of the user interface to utilise the faceted nature of both the thesauri and the user queries. A fixed set of views where the thesaurus provided the central view (Ref 14) placed too high a cognitive load on the user and did not lend itself to the flexible interaction sought by the designers. Much of the system state required by the user to make the system predictable (Ref 4) was hidden, and although the initial user interface provided means whereby the user could determine the set of values and terms stored in the system to be used in progressing the query, this set is typically too large to be held in the user's memory. The redesign locates the information needed to understand the behaviour of the system in the display (Ref 7).

4. HIBROWSE for EMBASE

The first demonstrable view-based searching systems provide enhanced access to the six million-plus record EMBASE database of biomedical bibliographic records. Assessing the usability of view-based techniques for EMBASE is a project, principally funded by the British Library Research and Innovation Centre, which involves collaboration between the database publisher, hardware and software providers, a database host organisation and a university centre for research and development. Multiple views using the EMTREE thesaurus of 38,000 terms can be accessed and combined with other views, such as year of publication, as required by the query.

In the following example interaction a user is searching for information on the therapy of Alzheimer Disease using the prototype HIBROWSE for EMBASE on the ADABAS system, which at the time contained close to 300,000 records.

4.1. Finding a view

The user selects the diseases facet from the drop down list of views available (Figure 1) and has the option either to browse from the top-level thesaurus terms in this facet or to find a view by keying the first characters from the term, or a synonym, and selecting a term from a permuted list. In this example the user types 'alz' and selects 'alzeimer disease' (Figure 2).

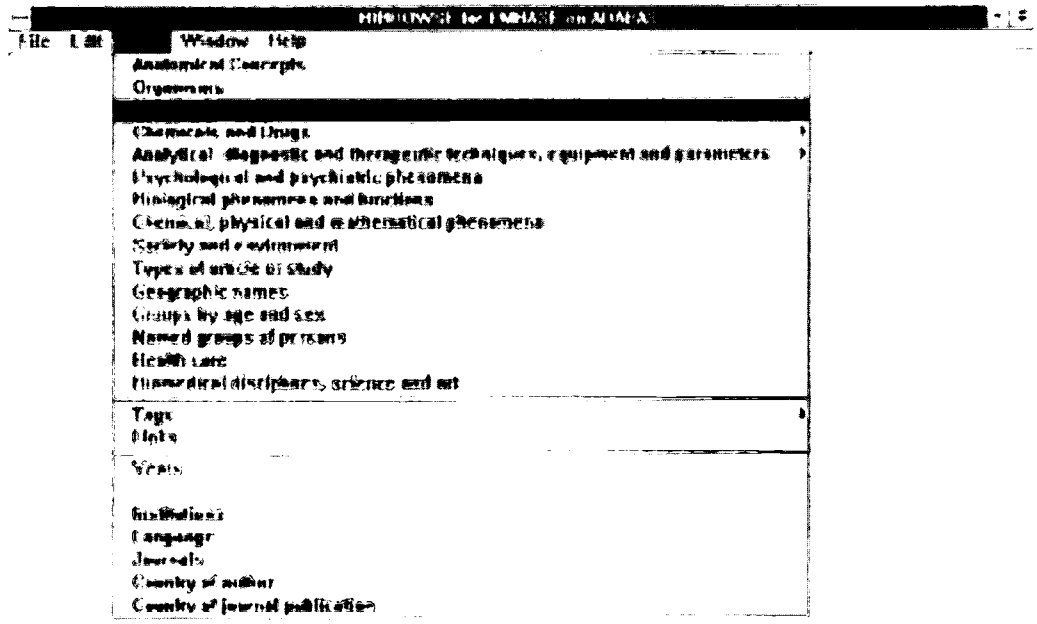

Figure 1: Selecting a facet.

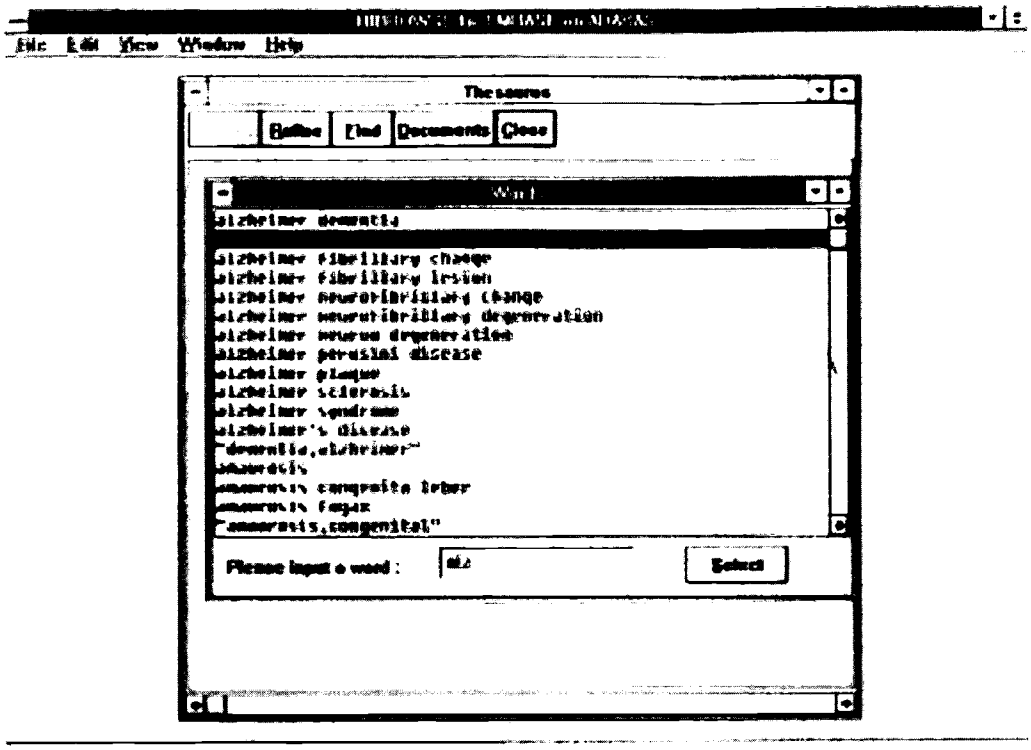

Figure 2: Using the permuted term look-up list.

The user is then presented with the view containing 'alzheimer disease': given that this term does not have any narrower terms in the EMTREE thesaurus it is presented in the context of the more general term 'degenerative disease' (Figure 3). There are 1297 records dealing with Alzeimer Disease in this subset of EMBASE.

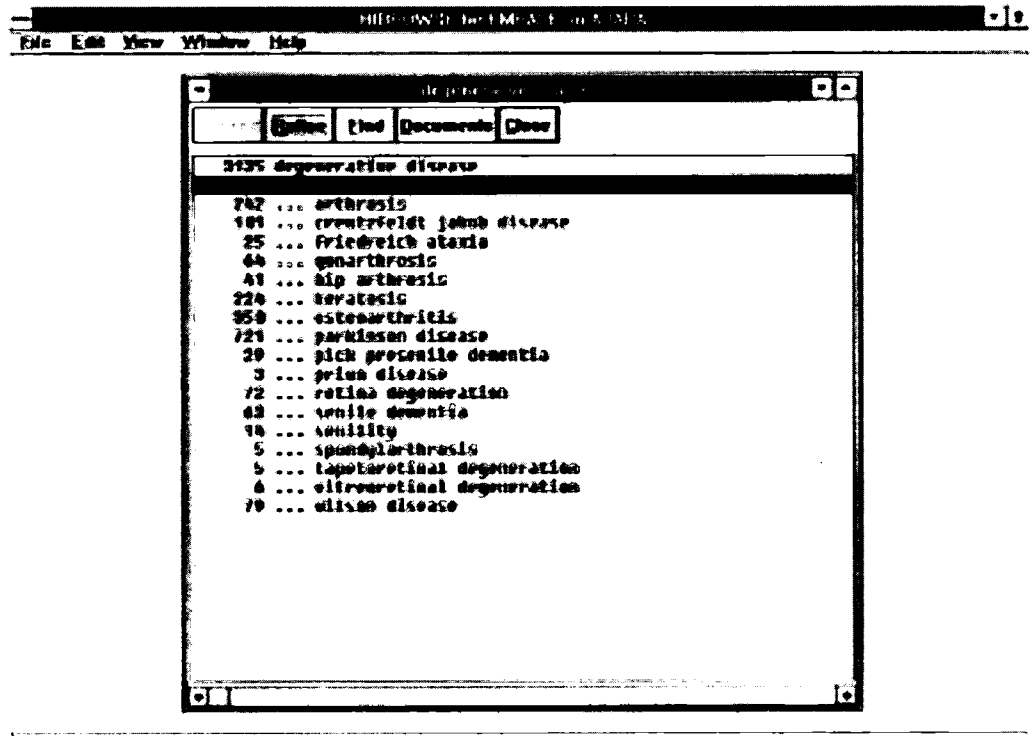

Figure 3: The view of degenerative disease.

4.2. Refining a view by adding a second view

The user wants to reduce the number of records and decides to refine the view in Figure 3 through the selection of a second view concerning 'therapy' from the drop-down list of facets. A choice is made to browse from the top level thesaurus terms (Figure 4).

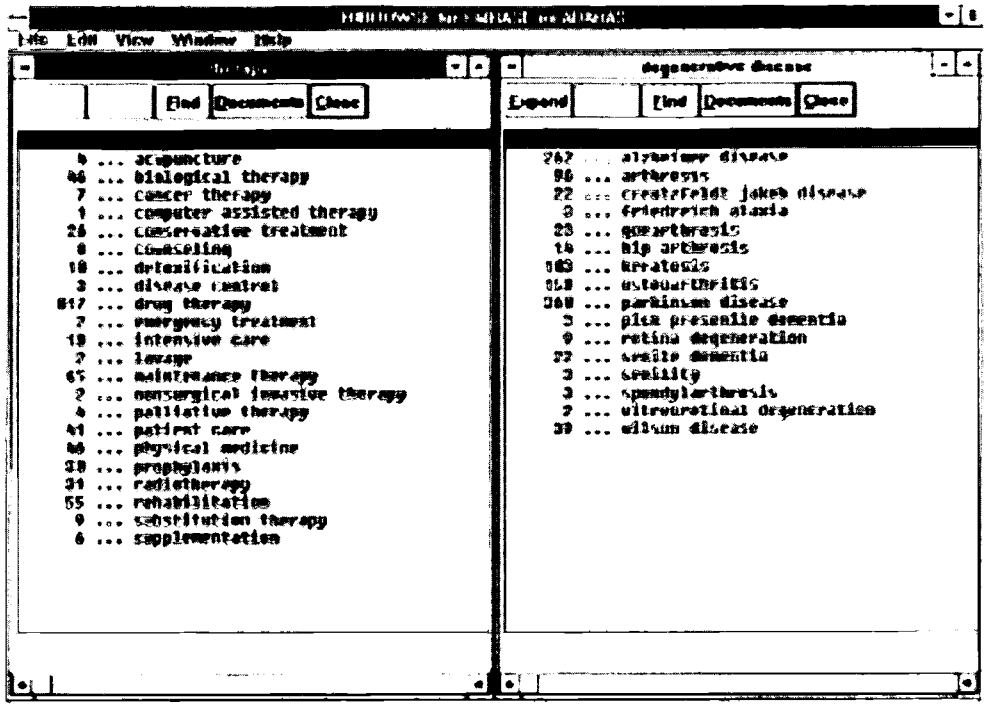

Figure 4: View refinement through the addition of a second view on therapy.

The number of records concerning 'alzheimer disease' has now reduced from 1297 to 262 as the second view on 'therapy' has acted as a filter onto the 'degenerative disease' view. Similarly the 'therapy' view itself has been refined to present only those documents which concern 'degenerative disease'. The original unfiltered view is presented in Figure 5.

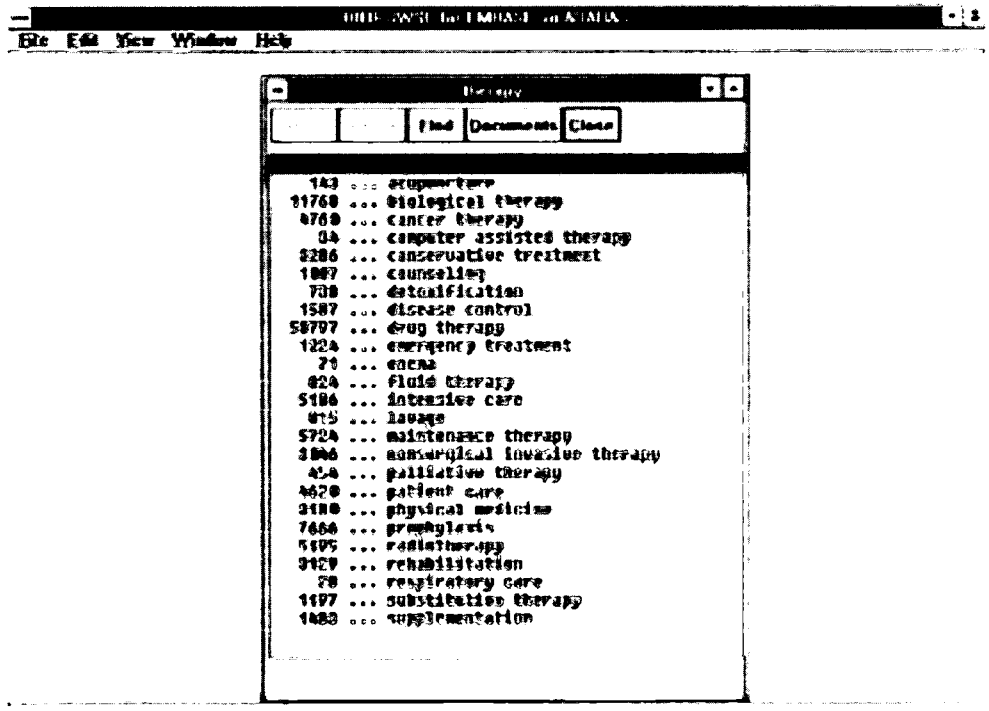

Figure 5: The unfiltered view of therapy.

4.3. Refining a view by selection from an existing view

Selecting 'drug therapy' on the new view, and electing to refine the views further, reduces the number of records being considered down from 262 to 190, although the numbers of records in the drug therapy view refer to the full set of 817 degenerative disease records (Figure 6).

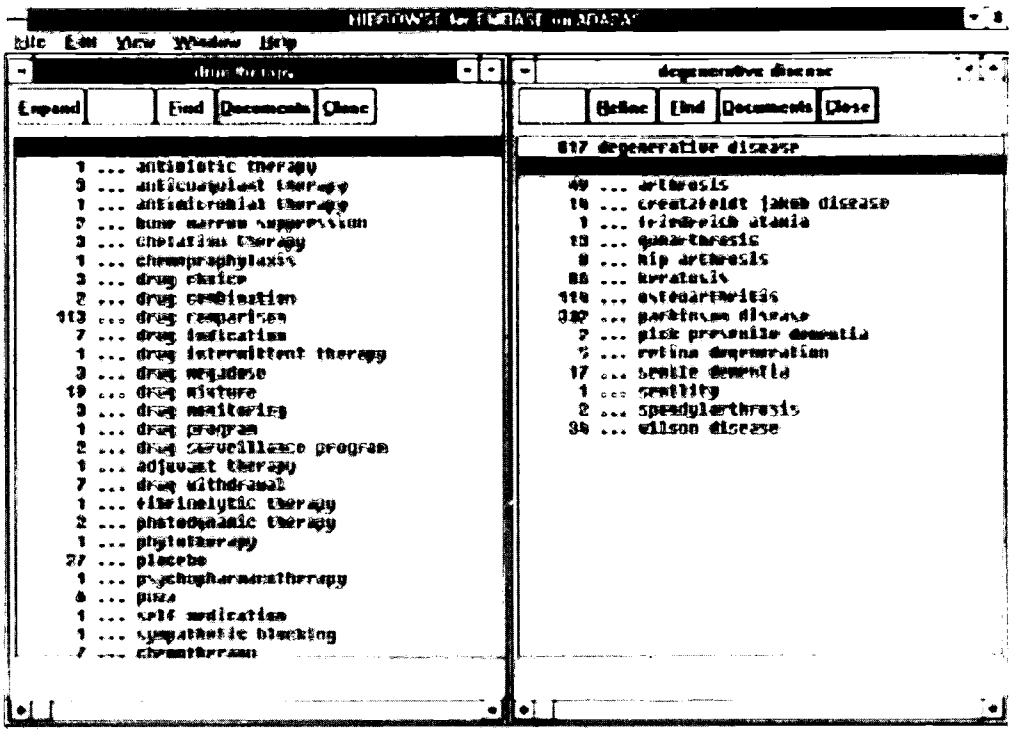

Figure 6: Refinement through selection within an existing view.

The final refinement comes about when the user refines the degenerative disease view to contain only 'alzheimer disease' (Figure 7). All the thesaurus terms from the drug therapy view which fail to reference any of the documents in the current set are removed from the refined therapy view and the user chooses to see the three titles of those documents concerning 'hormonal therapy'.

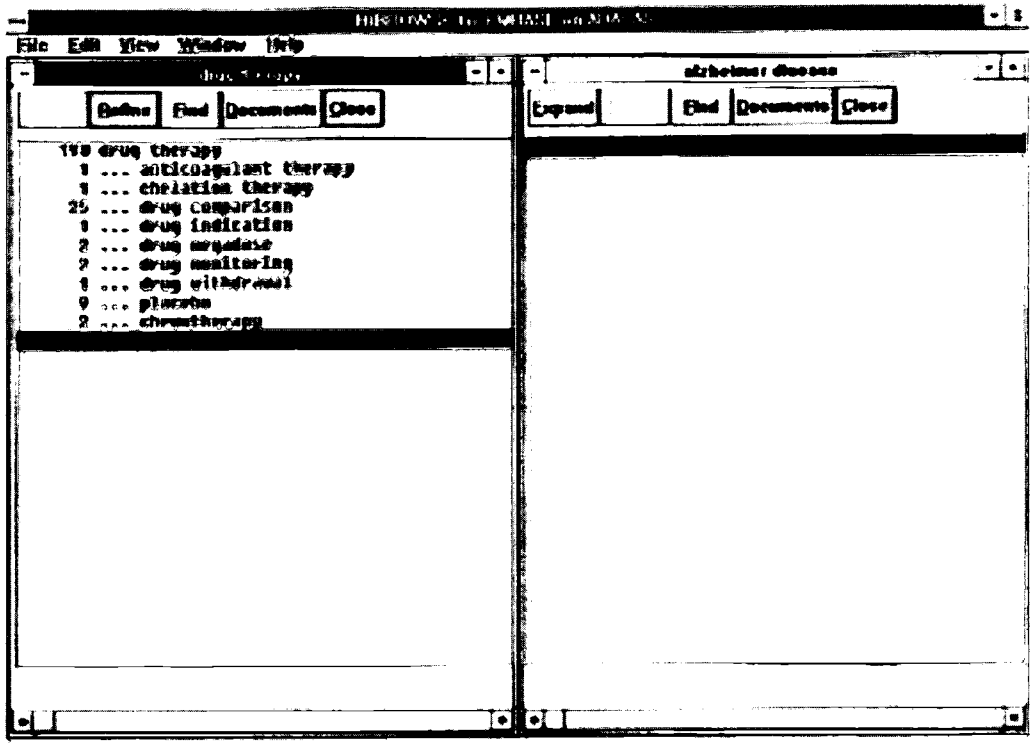

Figure 7: Final views of alzheimer disease and drug therapy.

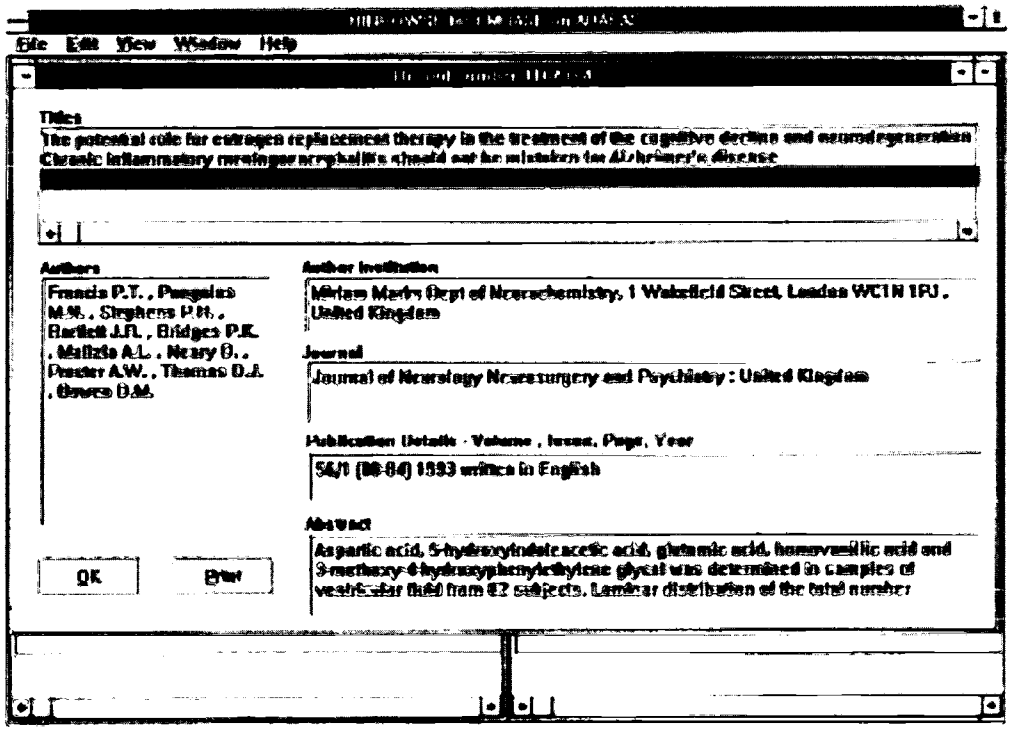

Figure 8: Records concerning Alzheimer disease and hormonal therapy.

BEST COPY AVAILABLE

4.4. Expanding a view

Should users find the number of records drops to near zero they are able to expand either or both of the views, relaxing the filtering constraints and returning additional terms and associated records to the views.

4.5. Additional views

Further developments will provide views onto the database via authors, author institutions, journals, language of article, country of publication and other attributes of the bibliographic record. The provision of such views will considerably enhance the functionality available to the user without a commensurate increase in complexity. Such features facilitate database analysis as well as high performance searching. Figure 9 shows the additional views provided from the country of journal and country of author.

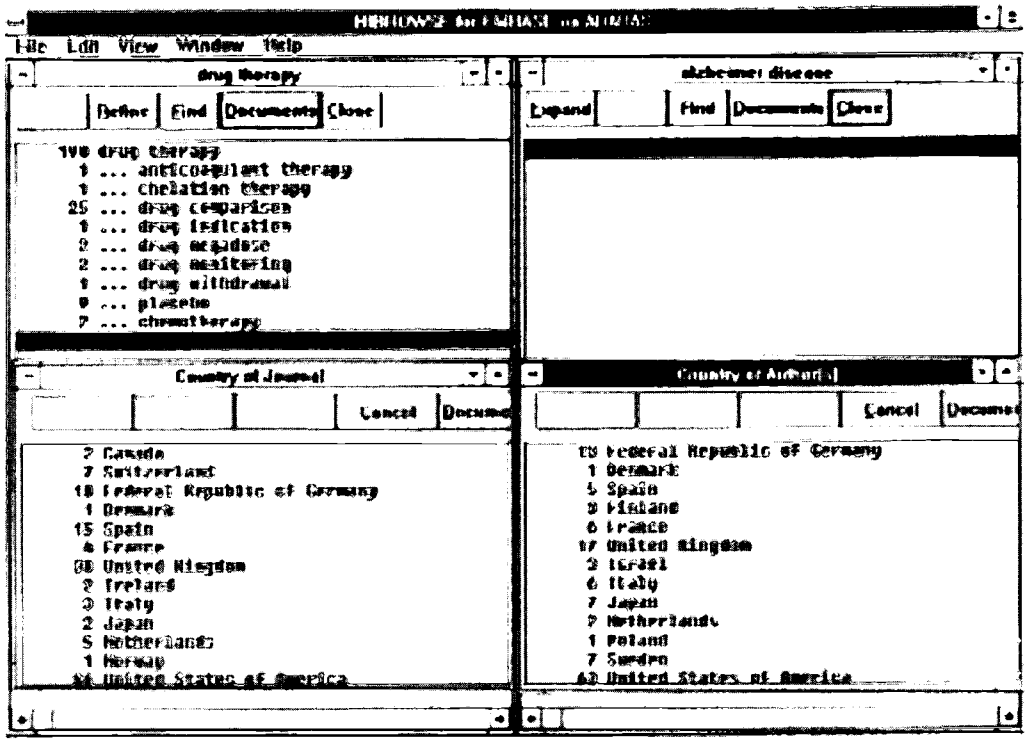

Figure 9: Adding the views Country of Journal and Country of Author.

5. HIBROWSE for EPOQUE

HIBROWSE for EPOQUE, the document database of the European Parliament, takes advantage of the multilingual EUROVOC thesaurus and demonstrates the global potential for improving access to databases where English is not the first language of the user. The following example illustrates a second prototype development which has explored the use of in-line expansion of thesaurus views and the use of additional icons to provide a more informative view. This client software development will also provide access to the full EMBASE database on the STN-International network.

The following example explores the subjects of 'Employment and work' and 'Business and competition'

5.1. Selecting a view and browsing through in-line expansion of concepts

The view on 'Employment and work' is updated online to the European Parliament where documents are added to the EPOQUE database on a daily basis (Figures 10, 11).

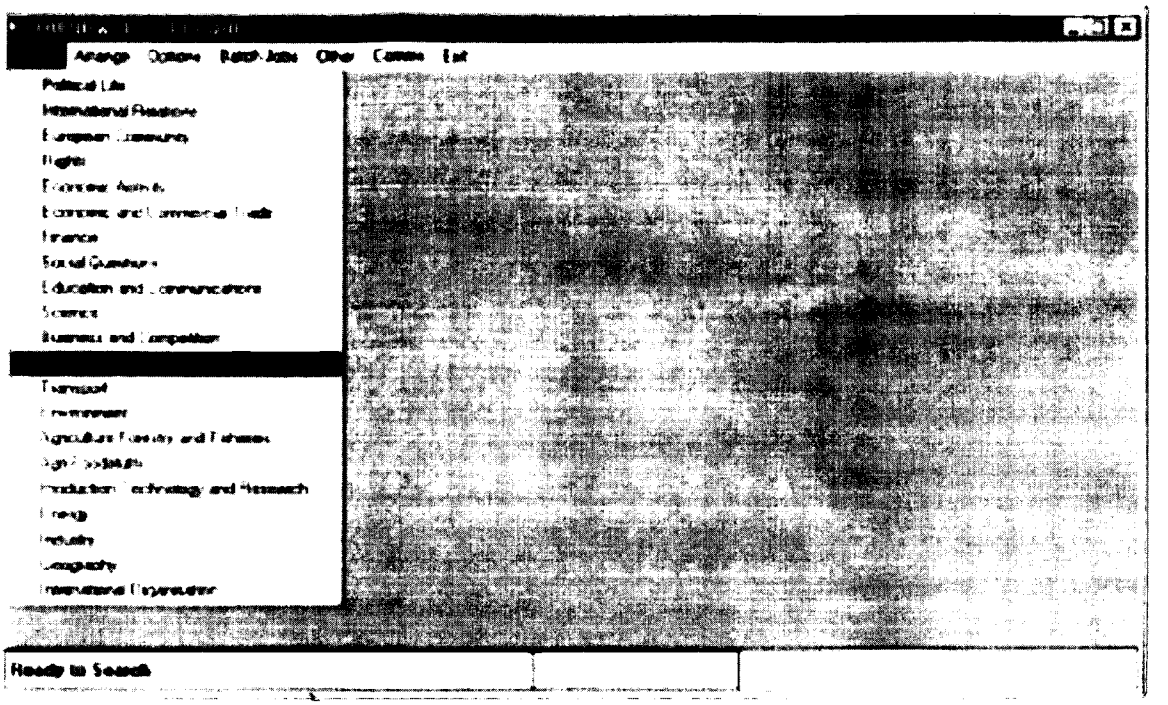

Figure 10: Selecting a view on HIBROWSE for EPOQUE.

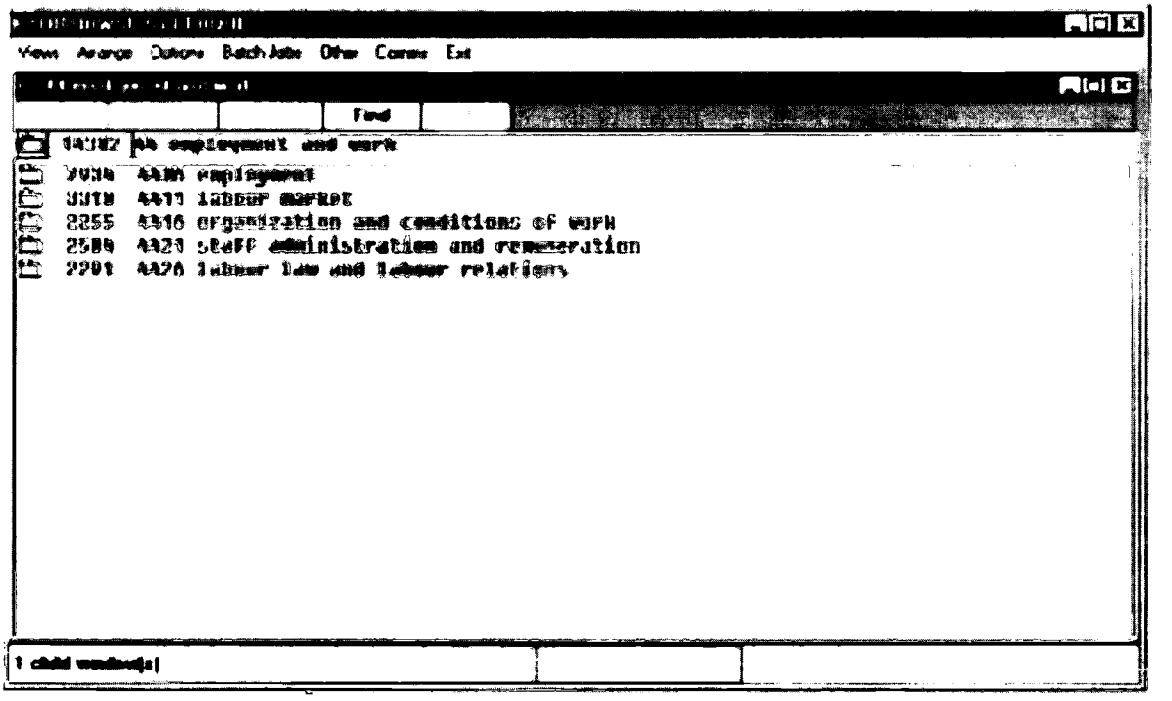

Figure 11: The top level view on 'Employment and work'.

The user can browse these documents by opening folders. Figure 12 shows how the view is expanded *in situ* revealing a breakdown of the 7934 documents into more specific headings.

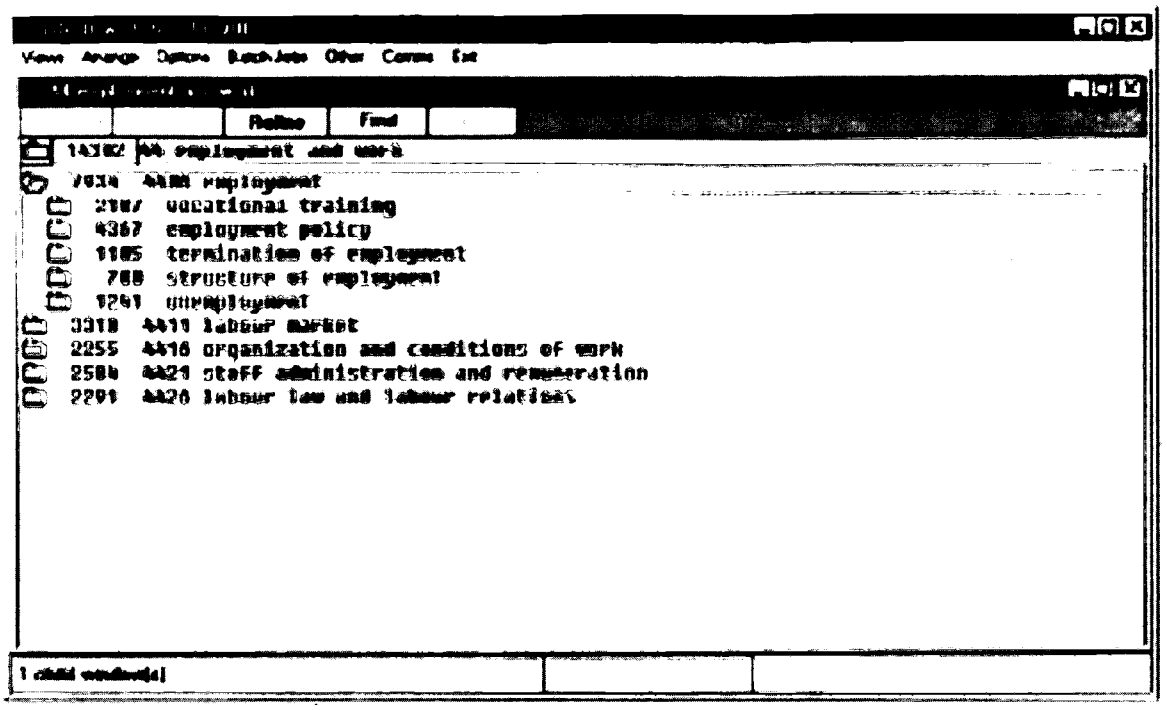

Figure 12: Expanding a view by opening a document folder.

5.2. Refinement through selection and the addition of views

This view is refined by the selection of a second view on 'Business and competition'. This reduces the number of documents from 14,302 to 2745. The user opens the folder to expand the business and competition view under 'types of undertaking' and then 'undertaking'. The user selects to refine by selecting all documents in 'undertaking' which are marked as seen in Figure 13.

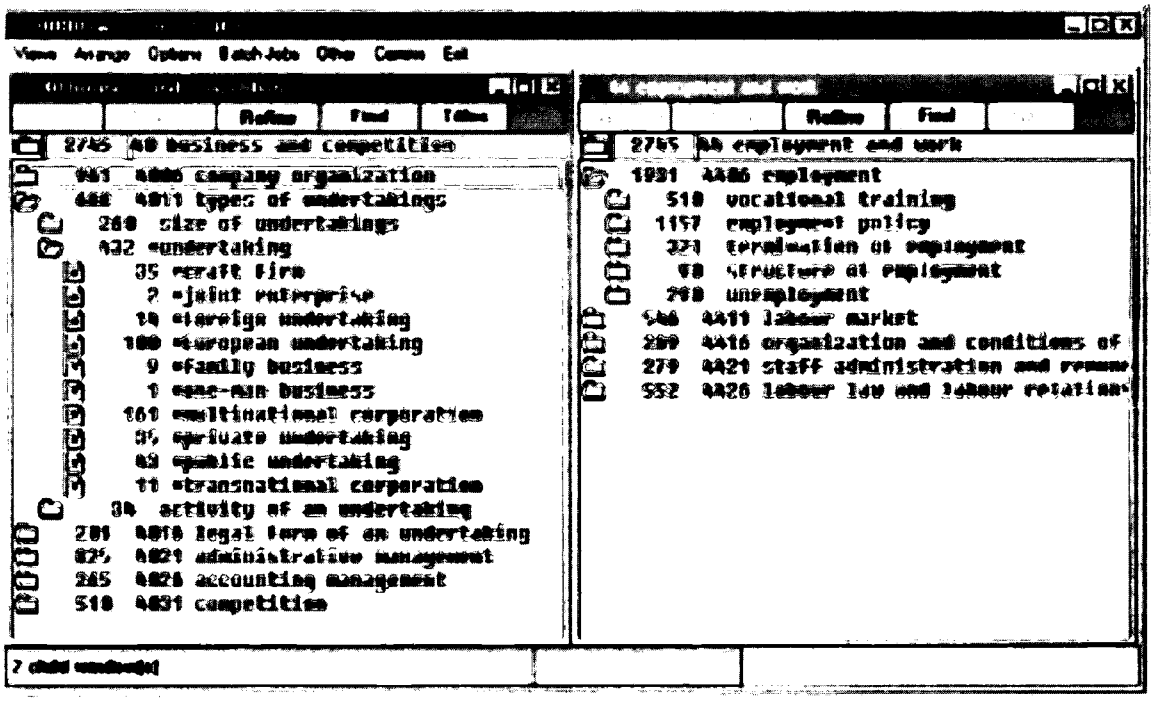

Figure 13: Combining 'Business and Competition' and 'Employment and Work'.

This clears the 'Business and competition' view of all except the selected documents and refines the view on 'Employment and work' as shown in Figure 14.

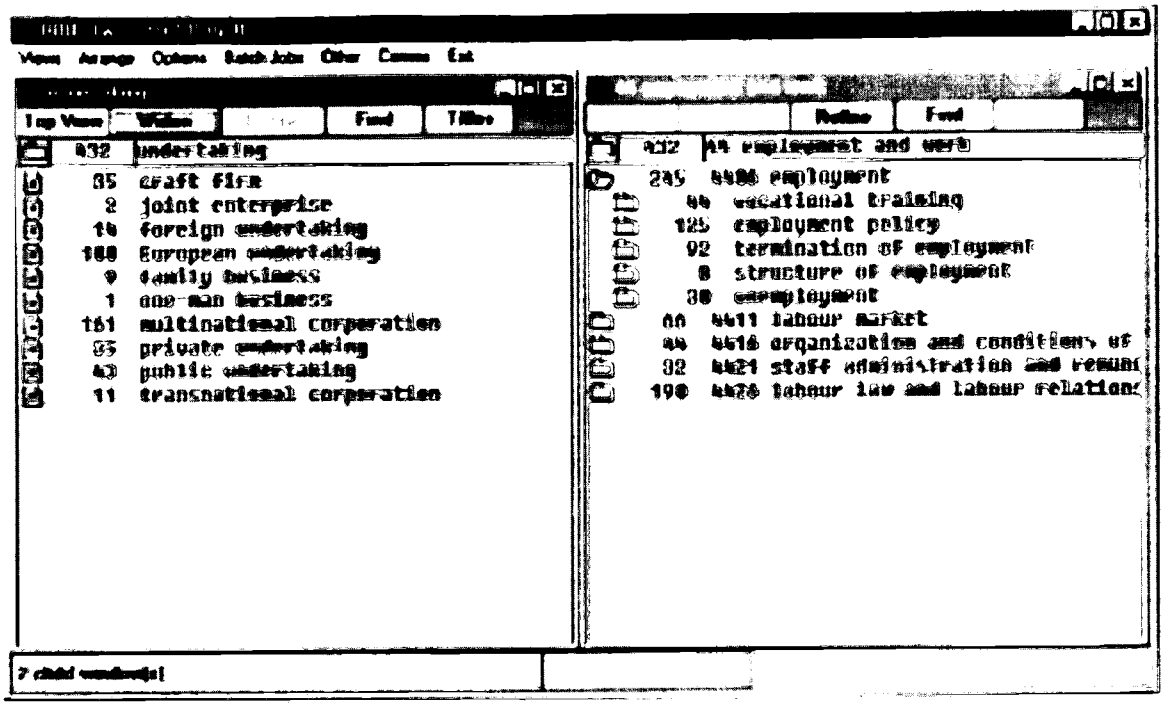

Figure 14: Refining 'Employment and work' by 'Undertaking'.

Additional refinements are pursued through the selection of labour law and labour relations and the addition of the top level industry view (Figure 15).

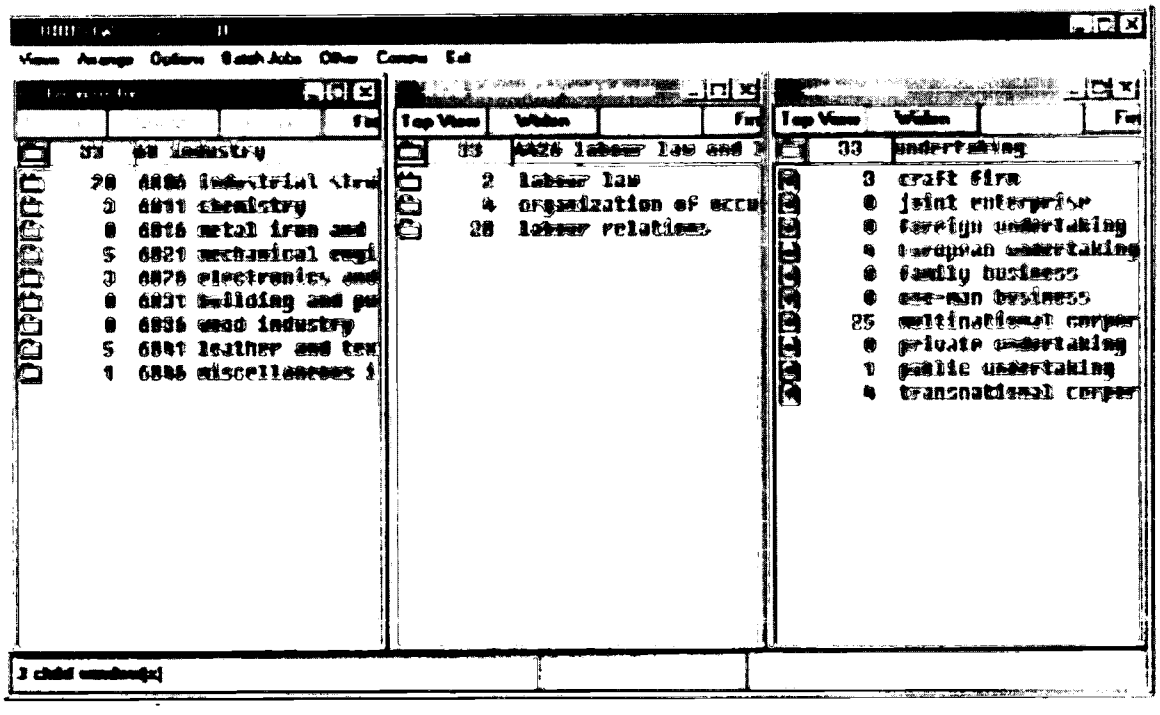

Figure 15: Additional refinement through selections and adding the view on Industry.

6. Conclusions

Effective disintermediation in information retrieval requires the development of tools and techniques that make it possible for the user to satisfy an information need without investing in the acquisition of skills and knowledge of the search process, currently the province of information scientists and librarians. The view-based searching techniques applied in the HIBROWSE prototype systems, described above, demonstrate how the incorporation of knowledge structures, primarily thesauri and classification schemes, in the user interface enables both detailed and general examination of the contents of a database and extends the scope for the user to control the interaction by adding, expanding, refining or removing views into the database. These systems also suggest how users can more effectively interact with a retrieval system when they are unable to specify the subject of their need, or can only express it in general terms and would like simply to explore the contents of the database.

The complexity and detail of indexing and retrieval operations can be hidden through the use of implicit Boolean searching, enacted whenever the perspective of a search is changed by an operation on just one of the views onto the database. Interaction with the system by the user is most concerned with the subject matter of the information need and the user is not concerned with the process of searching but rather with the specification of the subject matter of interest. Learning the capabilities of a system remains an essential for the professional intermediary through training courses, manuals and experience doing searching, but the capabilities of the system should ideally be self evident for the end-user and made more accessible, even if these are only required infrequently. The assumption that a sacrifice has to be made in performance where the interface is made more 'user-friendly' is being challenged through the application of techniques which emphasise the importance of recognition and selection and implicit searching as central to the design of online information retrieval interaction.

Where the view-based searching approach as implemented in HIBROWSE can be improved is the subject of on-going research; what is clear, by examining the workings of the current prototypes, is the increase in searching power made available without a commensurate increase in cognitive effort on the part of the user. Remembering commands and how to apply them is replaced by a limited set of actions which can reduce or expand a set of documents by refining or expanding, adding or removing views. The ease of use in World Wide Web navigation might now be complemented by an effectiveness in retrieval. Large result sets can be viewed according to helpful arrangements and divisions which indicate ways to limit or expand the amount of information being retrieved.

The absence of indexing using a controlled vocabulary for any information collection does not preclude the application of this approach. View-based techniques are applicable to both heavily indexed and free text databases alike, given there are suitable thesauri which, in the case of the latter, can be used to generate extended queries through the use of synonyms and narrower terms with stemming and other free text searching devices.

There will still be users who are content with the results of a 'quick and dirty' search, happy to use the five 'top ranked' documents from a set of many thousands, who will not reap the benefits of these efforts at disintermediation through the application of knowledge structures and associated search techniques. However, we trust there will be many others who, in order to exercise and extend their own professional skills and knowledge, will welcome the opportunity to discern the most apposite articles for the task they have in hand through the application of view-based searching.

Dr A, Steven Pollitt
Director, CeDAR — Centre for Database Access Research
School of Computing and Mathematics
University of Huddersfield
Queensgate
Huddersfield HD1 3DH
UK
Tel: +44 (0)1484 472147
Fax: +44 (0)1484 421106
E-mail: a.s.pollitt@hud.ac.uk

Acknowledgements

The HIBROWSE for Bibliographic Databases project and two of the authors are supported by a British Library Research and Innovation Centre grant RDD/G/282. Additional sponsorship from Sun Microsystems Inc., Software AG of the United Kingdom Limited, Elsevier Science BV, Elonex and the University of Huddersfield is gratefully acknowledged.

References

[1] CRG (1955) *The Need for a Faceted Classification as the Basis of all Methods of Information Retrieval.* A memorandum of the Classification Research Group, published as UNESCO document 320/5515, 26 May 1955.

[2] Davies, R. (1996) Thesauri for the new century, *Macrothesaurus* no.4, May, OECD Development Centre, Paris.

[3] de Grolier, E. (1965) Current trends in theory and practice of classification, *Classification Research — Proceedings of the Second International Study Conference*, Elsinore, Denmark 14-18 September 1964, FID/CR Copenhagen.

[4] Dix, A.J. (1991) *Formal Methods for Interactive Systems*, Academic Press, London.

[5] Nicholas, D. (1995) Are information professionals really better online searchers than end-users? (And whose story do you believe?) *Proceedings of the 19th International Online Information Meeting* London, December 1995, Learned Information,, pp. 383–397.

[6] Occhini, G. (1983) Microelectronics and teleinformatics as economic and social disintermediation tools. (Elimination of intermediaries between producers, distributors and consumers), *BIAS Microelettronica. International Conference on Microelectronics and Teleinformatics*, Milan, Italy 23–25 February 1983, Federazioone delle Associazioni Scientifiche e Techniche, p. 9.

[7] Payne, S.J. (1991) Display-based action at the user interface, *International Journal of Man-Machine Studies*, **35**, 275–289.

[8] Pollitt, A.S. (1981) An expert system as an online search intermediarym *Proceedings of the 5th International Online Information Meeting*, London, December 1981, Learned Information, pp. 25–32.

[9] Pollitt, A.S. (1985) End-user Boolean searching on viewdata using numeric keypads, *Proceedings of the 9th International Online Information Meeting*, London, December 1985, Learned Information, pp. 373–380.

[10] Pollitt, A.S. (1987) CANSEARCH: An expert systems approach to document retrieval, *Information Processing and Management*, **23**(2), 119–138.

[11] Pollitt, A.S. (1988) A common query interface using MenUSE — A Menu-based User Search Engine, *Proceedings of the 12th International Online Information Meeting*, London, December 1988, Learned Information, Vol. 2, pp. 445–457.

[12] Pollitt, A.S., G.P. Ellis, M.P. Smith, M.R. Gregory, C.S. Li and H. Zangenberg (1993) A common query interface for multilingual document retrieval from databases of the European Community Institutions, *Proceedings of the 17th International Online Information Meeting*, London, December 1993, Learned Information, pp. 47–61.

[13] Pollitt, A.S., G.P. Ellis, M.P. Smith (1994) Peek-a-Boo revived — end-user searching of bibliographic databases using filtering views, *Proceedings of the 18th International Online Information Meeting*, London, December 1994, Learned Information, pp. 63-74.

[14] Pollitt, A.S., G.P. Ellis, M.P. Smith (1994) HIBROWSE for bibliographic databases, *Journal of Information Science*, **20**(6), 413–426.

[15] Ranganathan, S.R. (1965) Library classification through the century, *Classification Research — Proceedings of the Second International Study Conference*, Elsinore, Denmark, 14–18 September 1964, FID/CR Copenhagen.

[16] Treglown, M. (1996) Results of applying the cognitive walkthrough method to the HIBROWSE for EMBASE system, *Research Report RR96/02*, School of Computing and Mathematics, University of Huddersfield.

[17] Treglown, M. (to appear) Are database thesauri menus? To appear in the *Adjunct Proceedings of HCI '96*, held London, 20–23 August 1996.

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement (OERI)
Educational Resources Information Center (ERIC)

NOTICE

REPRODUCTION BASIS

☒ This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

☐ This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").